DOCUMENT RESUME

ED 415 280 TM 028 015

AUTHOR House, J. Daniel; Gupta, Suman; Xiao, Beiling
TITLE Gender Differences in Prediction of Grade Performance from Graduate Record Examination Scores.
PUB DATE 1997-11-00
NOTE 17p.; Paper presented at the Annual Meeting of the Illinois Association for Institutional Research (Chicago, IL, November 6-7, 1997).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *College Entrance Examinations; Grades (Scholastic); *Graduate Students; Graduate Study; Higher Education; Performance Factors; *Prediction; Scores; *Sex Differences; Test Validity
IDENTIFIERS *Graduate Record Examinations

ABSTRACT

One focus of educational measurement has been the examination of the relationship between students' entering academic characteristics and their subsequent achievement outcomes. In the case of admissions tests, an important consideration is that test scores predict later achievement similarly for all groups of students who take the test. The purposes of this study were to: (1) assess the validity of the Graduate Record Examination (GRE) for predicting students' grade performance; and (2) investigate gender bias in the prediction of student performance from GRE scores. A sample of 6,015 graduate students in 8 academic disciplines was examined. GRE scores were found to be significant predictors of subsequent grade performance for students in several academic disciplines. In addition, the findings from this study indicate that, in some instances, admissions test scores may not predict later achievement similarly for male and female students. (Contains 3 tables and 36 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 415 280

Gender Differences in Prediction of Grade Performance From Graduate Record Examination Scores

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Daniel House.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

J. Daniel House

Suman Gupta

Beiling Xiao

Northern Illinois University

TM028015

Paper presented at the Illinois Association for Institutional Research Annual Meeting, Chicago, Illinois, November 6-7, 1997

BEST COPY AVAILABLE

Abstract

One focus of educational measurement has been the examination of the relationship between students' entering academic characteristics and their subsequent achievement outcomes. In the case of admissions tests, an important consideration is that test scores predict later achievement similarly for all groups of students who take the test. The purposes of this study were to: (a) assess the validity of the Graduate Record Examination for predicting students' grade performance, and (b) to investigate gender bias in the prediction of student performance from GRE scores. A sample of 6,015 graduate students in eight academic disciplines was examined in this study. GRE scores were found to be significant predictors of subsequent grade performance for students in several academic disciplines. In addition, the findings from this study indicate that, in some instances, admissions test scores may not predict later achievement similarly for male and female students.

One focus of higher education assessment has been an examination of the relationship between students' entering academic characteristics and their subsequent achievement outcomes. For prospective graduate students, one factor that is given consideration during the admissions process is performance on the Graduate Record Examination (GRE). A number of studies have assessed the validity of the GRE for predicting students' later grade performance. For example, GRE scores have been found to be significant predictors of grade performance in specific courses (Goldberg & Alliger, 1992; House, Johnson, & Tolone, 1987; Huitema & Stein, 1993) and of cumulative grade performance (Federici & Schuerger, 1974; House & Johnson, 1993b) for graduate students in psychology. Considering graduate students in education, GRE scores have been shown to be significantly correlated with grades in specific courses such as statistics and research methods (Furst & Roelfs, 1979) and with overall grade performance (Camp & Clawson, 1979; House, 1989; Omizo & Michael, 1979). However, GRE scores showed significant (although low) correlations with later grade performance for students in a Graduate School of Social Work (Milner, McNeil, & King, 1984). Oldfield and Hutchinson (1997) found significant correlations between GRE scores and grades in specific courses (statistics and research methods) for students in a Master of Public Administration program. Further, recent research has assessed the predictive validity of the GRE for minority (American Indian/Alaska Native) students (House, 1997) and for outcomes such as degree completion (Holmes & Beishline, 1996; House & Johnson, 1992, 1993a) or preliminary examinations performance (Dollinger, 1989). It has been noted that the validity of the GRE may vary considerably by major field of study (Thornell & McCoy,

1985) and that further research is needed to assess the validity of the GRE for a number of academic disciplines.

A critical issue in the use of admissions test scores for selection purposes is that test scores predict subsequent achievement similarly for all groups of students who take the test. Previous studies have found that, in some instances, admissions test scores do not predict achievement outcomes equally well for male and female students. In general, test scores tend to underpredict the academic performance of female students and to overpredict the grade performance of male students. In other words, female students tend to earn grades that are higher than are predicted by their test scores while male students tend to earn grades that are lower than predicted by their test scores. Considering undergraduate students, research has indicated that both SAT and ACT scores tend to underpredict female students' GPA (Breland & Griswold, 1982; Linn, 1973; Sawyer, 1986). Recent research has indicated that course characteristics may explain some of the gender differences in achievement (Keller, Crouse, & Trusheim, 1993) while factors such as academic preparation and study behaviors may be related to the underprediction of female students' performance from their test scores (Stricker, Rock, & Burton, 1993). Finally, gender bias has been found for the prediction of graduate grade performance from Miller Analogies Test scores (House & Keeley, 1993, 1995).

There has been relatively little research on gender bias in the prediction of graduate grade performance from GRE scores. Kaczmarek and Franco (1986) found that GRE scores did not predict subsequent grade performance similarly for a small sample of male and female

students in a master's-level program in counseling. House (1994) found that GRE scores predicted grade performanee significantly more strongly for female students than for male students for a large sample of students in education. There is a need, however, for research to assess gender bias in the prediction of grade performance from GRE scores for students in a variety of academic disciplines.

There were two purposes of this study. First, this study was designed to assess the validity of the GRE for predicting subsequent grade performance for students in several academic disciplines. Second, this study was intended to investigate gender bias in the prediction of student performance (graduate GPA) from GRE scores.

Methods

Students

The sample consisted of 6,015 students enrolled at a public university during a nine-year period. Students who had completed at least six semester hours of graduate credits were included in the sample. Students in eight disciplines were assessed: Chemistry, Biological Sciences, English, History, Education, Psychology, Communicative Disorders, and Economics. Variables that were collected for each student included gender, cumulative graduate GPA, and GRE scores: Verbal (GRE-V), Quantitative (GRE-Q), and GRE-Total (verbal + quantitative).

Methods

The data from this study were analyzed in two ways. First, validity coefficients were computed for the relationship between each GRE score and graduate GPA for the entire

sample, and separately for male and female students. The correlation coefficients obtained for the men and women were then compared for equivalence using Fishers's Z-transformation test (Kleinbaum, Kupper, & Muller, 1988); this procedure has been used to compare the validity of the GRE for subgroups of students (House, 1994). Second, the method used to investigation prediction bias was to compare students by gender group on the mean error of prediction found for the GRE. A least-squares regression equation based on data from all students was used to compute a predicted GPA for each student, and a residual score (the predicted GPA minus the actual GPA) was then computed for each individual. The mean error for each group was the mean of the residual scores. The gender groups were then compared using analysis of variance to test for systematic error in the prediction of graduate GPA from GRE scores (Reynolds, 1982). These procedures were used for each test score (GRE-V, GRE-Q, and GRE-Total) and for each of the eight disciplines included in this study.

Results

Descriptive statistics for GRE scores and GPA are summarized in Table 1; data are provided for all students in each academic discipline as well as separately for male and female students. Correlations between each section of the GRE and grade performance for students in each program are shown in Table 2. Several significant findings were obtained. Considering GRE-V scores, significant correlations were found for five programs (Biology, English, History, Communicative Disorders, and Education) for the entire sample. Similarly, GRE-Q scores were significant predictors of GPA for students in six programs

(Biology, English, History, Communicative Disorders, Education, and Economics). Finally, GRE-Total scores were significantly correlated with GPA for students in five academic disciplines (Biology, English, Education, History, and Communicative Disorders). Further, no significant differences between male and female students were noted for any of the relationships between GRE scores and GPA for any of the academic programs included in this study.

Findings from the analyses of gender differences in the prediction of graduate GPA from GRE scores are presented in Table 3. Considering GRE-Q scores, two significant findings were obtained. For students in Education, there was a significant difference between male and female students in the prediction of graduate GPA from GRE-Q scores, $F(1,4583) = 101.13$, $p = .0001$; similar findings were also noted for students in psychology, $F(1,217) = 14.85$, $p = .0002$). In both instances, GRE-Q scores underpredicted the grade performance of female students. Considering GRE-V scores, significant differences between male and female students in the prediction of GPA from test scores were obtained for students in Education ($F(1,4583) = 66.89$, $p = .0001$) and in psychology ($F(1,217) = 11.23$, $p = .0010$). As before, GRE-V scores underpredicted the grade performance of female students. Finally, significant differences for GRE-Total scores were also seen for students in Education ($F(1,4583) = 102.54$, $p = .0001$) and in psychology ($F(1,217) = 12.80$, $p = .0004$). In each case, GRE scores underpredicted the GPA of female students and overpredicted the GPA of male students. In other words, female students earned grades that were higher than were predicted by their GRE scores while male students earned graduate grades that were

lower than had been predicted by their GRE scores.

Discussion

The results of this study indicate that, for some academic programs, GRE scores did not predict subsequent graduate grade performance similarly for male and female students. The pattern of underprediction of female students' GPA is consistent with findings previously seen for undergraduate admissions tests (McCornack & McLeod, 1988), the Miller Analogies Test (House & Keeley, 1993), and the GRE (House, 1994). These findings suggest that female students tend to earn higher grades than were predicted by their admissions test scores. The results of this study also indicate that GRE scores were significantly correlated with later grade performance for students in several academic disciplines. This pattern is consistent with results from previous studies of the validity of the GRE (Dollinger, 1989; House, 1983).

There are some limitations to the present study. First, students at only one institution were included in this study. Further research is needed to determine if these results would be found for students at other types of institutions. Previous multi-institution studies have been conducted for the ACT (House & Keeley, 1997), the Law School Aptitude Test (Rubin, 1980), and the Graduate Management Admission Test (Zwick, 1993). A second limitation of this study is that no analysis of the effects of restriction of range was made. It has been shown that restricting the distribution of admissions test scores for students who have been admitted for graduate study will produce lower than expected validity coefficients (Givner & Hynes, 1979). Finally, other research is needed to determine if these findings would be

replicated for other types of graduate outcomes such as degree completion or grades in specific courses (Hartnett & Willingham, 1980).

Institutional researchers are often requested to assess the validity of admissions test scores on their campuses. These results demonstrate that: (a) test scores may not predict later grade achievement similarly for all groups of students, and (b) provide an introduction to statistical methods used for the assessment of prediction bias. These results have implications for institutional researchers conducting assessment studies and for practitioners utilizing the results of those validity assessments (Astin, 1995).

References

Astin, A.W. (1995). Introduction to the IEO model and the College Student Survey (CSS). In A.W. Astin, E.L. Dey, W.S. Korn, & L.J. Sax (Eds.), *Analyzing CIRP Data: A Hands-On Workshop*. Los Angeles, CA: Higher Education Research Institute and UCLA Graduate School of Education.

Breland, H.M., & Griswold, P.A. (1982). Use of a performance test as a criterion in a differential validity study. *Journal of Educational Psychology, 74,* 713-721.

Camp, J., & Clawson, T. (1979). The relationship between the Graduate Record Examinations aptitude test and graduate grade point average in a Master of Arts counseling program. *Educational and Psychological Measurement, 39,* 429-431.

Dollinger, S.J. (1989). Predictive validity of the Graduate Record Examination in a clinical psychology program. *Professional Psychology: Research and Practice, 20,* 56-58.

Federici, L., & Schuerger, J. (1974). Prediction of success in an applied M.A. psychology program. *Educational and Psychological Measurement, 34,* 945-952.

Furst, E.J., & Roelfs, P.J. (1979). Validation of the Graduate Record Examinations and the Miller Analogies Test in a doctoral program in education. *Educational and Psychological Measurement, 39,* 147-151.

Givner, N., & Hynes, K. (1979). Achievement test validity: Correcting for restriction effects. *College and University, 54,* 119-123.

Goldberg, E.L., & Alliger, G.M. (1992). Assessing the validity of the GRE for students in psychology: A validity generalization approach. *Educational and Psychological Measurement, 52,* 1019-1027.

Hartnett, R.T., & Willingham, W.W. (1980). The criterion problem: What measure of success in graduate education? *Applied Psychological Measurement, 4,* 281-291.

Holmes, C.B., & Beishline, M.J. (1996). Correct classification, false positives, and false negatives in predicting completion of the Ph.D. from GRE scores. *Psychological Reports, 79,* 939-945.

House, J.D. (1983). Effects of restriction of range on predictive validity for the Graduate Record Examination. *Psychological Reports, 53,* 710.

House, J.D. (1989). Age bias in prediction of graduate grade point average from Graduate Record Examination scores. *Educational and Psychological Measurement, 49,* 663-666.

House, J.D. (1994). Gender differences in prediction of grade performance from Graduate Record Examination scores. *Journal of Psychology, 128,* 695-697.

House, J.D. (1997). Predictive validity of Graduate Record Examination scores for outcomes of American Indian/Alaska Native students. *Psychological Reports, 81,* 337-338.

House, J.D., & Johnson, J.J. (1992). Predictive validity of Graduate Record Examination scores and undergraduate grades for length of time to completion of degree. *Psychological Reports, 71,* 1019-1022.

House, J.D., & Johnson, J.J. (1993a). Graduate Record Examination scores and academic background variables as predictors of graduate degree completion. *Educational and Psychological Measurement, 53,* 551-556.

House, J.D., & Johnson, J.J. (1993b). Predictive validity of the Graduate Record Examination Advanced Psychology Test for graduate grades. *Psychological Reports, 73,* 184-186.

House, J.D., Johnson, J.J., & Tolone, W.L. (1987). Predictive validity of the Graduate Record Examination for performance in selected graduate psychology courses. *Psychological Reports, 60,* 107-110.

House, J.D., & Keeley, E.J. (1993). *Differential prediction of graduate student achievement from Miller Analogies Test scores.* Paper presented at the Illinois Association for Institutional Research annual meeting, Oakbrook Terrace, Illinois. (ERIC Document Reproduction Service No. ED 364 605).

House, J.D., & Keeley, E.J. (1995). Gender bias in the prediction of graduate grade performance from Miller Analogies Test scores. *Journal of Psychology, 129,* 353-355.

House, J.D., & Keeley, E.J. (1997). Predictive validity of college admissions test scores for American Indian students. *Journal of Psychology, 131,* 572-574.

Huitema, B.E., & Stein, C.R. (1993). Validity of the GRE without restriction of range. *Psychological Reports, 72,* 123-127.

Kaczmarek, M., & Franco, J.N. (1986). Sex differences in prediction of academic performance by the Graduate Record Examination. *Psychological Reports, 59,* 1197-1198.

Keller, D., Crouse, J., & Trusheim, D. (1993). Relationships among gender differences in freshmen course grades and course characteristics. *Journal of Educational Psychology, 85,* 702-709.

Kleinbaum, D.G., Kupper, L.L., & Muller, K.E. (1988). *Applied Regression Analysis and Other Multivariable Methods (Second Edition).* Belmont, CA: Duxbury Press.

Linn, R.L. (1973). Fair test use in selection. *Review of Educational Research, 43,* 139-161.

McCornack, R.L., & McLeod, M.M. (1988). Gender bias in the prediction of college course performance. *Journal of Educational Measurement, 25,* 321-331.

Milner, M., McNeil, J.S., & King, S.W. (1984). The GRE: A question of validity in predicting performance in professional schools of social work. *Educational and Psychological Measurement, 44,* 945-950.

Oldfield, K., & Hutchinson, J.R. (1997). Predictive validity of the Graduate Record Examination with and without range restraints. *Psychological Reports, 81,* 211-220.

Omizo, M.M., & Michael, W.B. (1979). The prediction of performance in a counselor education master's degree program. *Educational and Psychological Measurement, 28,* 433-436.

Reynolds, C.R. (1982). Methods for detecting construct and predictive bias. In R.A. Berk (Ed.), *Handbook of Methods for Detecting Test Bias* (pp. 199-227). Baltimore, MD: Johns Hopkins University Press.

Rubin, D.B. (1980). Using empirical Bayes techniques in the law school validity studies. *Journal of the American Statistical Association, 75,* 801-816.

Sawyer, R. (1986). Using demographic subgroup and dummy variable equations to predict college freshmen grade average. *Journal of Educational Measurement, 23,* 131-145.

Stricker, L.J., Rock, D.A., & Burton, N.W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology, 85,* 710-718.

Thornell, J.G., & McCoy, A. (1985). The predictive validity of the Graduate Record Examinations for subgroups of students in different academic disciplines. *Educational and Psychological Measurement, 45,* 415-419.

Zwick, R. (1993). The validity of the GMAT for the prediction of grades in doctoral study in business and management: An empirical Bayes approach. *Journal of Educational Statistics, 18,* 91-107.

Table 1
Descriptive Statistics for GRE Scores and Cumulative GPA by Gender

		All			Males			Females		
Variable	Department	*n*	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
GRE-V&Q										
	Chemistry	152	1079.9	149.0	99	1072.6	147.3	53	1093.6	152.6
	Biology	153	1130.7	153.2	73	1123.6	164.1	80	1137.3	143.3
	English	367	1077.4	207.1	148	1086.3	218.7	219	1071.5	199.2
	History	173	1081.2	183.1	110	1080.5	161.7	63	1082.4	216.8
	Education	4,585	981.7	185.9	1,474	1009.4	189.6	3,111	968.6	182.8
	Psychology	219	1189.3	127.0	78	1189.7	119.8	141	1189.0	131.3
	Economics	137	1087.2	175.7	103	1070.9	173.1	34	1136.8	176.9
	Comm Disorders	229	969.6	137.1	13	971.5	173.1	216	969.5	135.2
GRE-V										
	Chemistry	152	430.1	115.1	99	423.2	112.1	53	443.0	120.5
	Biology	153	526.8	104.8	73	510.0	107.2	80	542.1	100.7
	English	367	566.7	118.8	148	560.6	119.5	219	570.9	118.5
	History	173	565.4	104.2	110	563.8	95.5	63	568.1	118.8
	Education	4,585	483.9	103.0	1,474	484.9	103.3	3,111	483.4	102.8
	Psychology	219	583.2	87.6	78	566.4	80.3	141	592.4	90.3
	Economics	137	438.1	123.8	103	419.7	108.6	34	493.8	149.8
	Comm Disorders	229	473.5	78.3	13	480.0	93.5	216	473.1	77.5
GRE-Q										
	Chemistry	152	649.8	94.9	99	649.4	94.8	53	650.6	96.0
	Biology	153	603.9	90.3	73	613.6	96.4	80	595.1	83.9
	English	367	510.7	123.7	148	525.7	125.6	219	500.6	121.7
	History	173	515.8	112.3	110	516.6	105.6	63	514.3	124.1
	Education	4,585	497.8	114.5	1,474	524.5	117.4	3,111	485.1	110.9
	Psychology	219	606.1	77.6	78	623.3	75.6	141	596.6	77.3
	Economics	137	649.1	175.7	103	651.2	109.1	34	642.9	84.7
	Comm Disorders	229	496.1	89.2	13	491.5	94.6	216	496.3	89.0
Cumulative GPA										
	Chemistry	152	3.45	0.428	99	3.43	0.426	53	3.49	0.435
	Biology	153	3.61	0.45	73	3.56	0.50	80	3.66	0.39
	English	367	3.59	0.38	148	3.57	0.38	219	3.60	0.38
	History	173	3.73	0.31	110	3.70	0.33	63	3.78	0.26
	Education	4,585	3.67	0.33	1,474	3.61	0.34	3,111	3.69	0.32
	Psychology	219	3.54	0.44	78	3.40	0.51	141	3.61	0.37
	Economics	137	3.26	0.56	103	3.24	0.58	34	3.31	0.52
	Comm Disorders	229	3.61	0.36	13	3.53	0.32	216	3.61	0.36

Table 2

Correlations Between GRE Scores and Cumulative GPA

GRE Test	Department	All	Males	Females	Z
GRE-V&Q					
	Chemistry	-.010	-.028	.010	0.14
	Biology	.283**	.297*	.259*	0.17
	English	.270**	.217**	.314**	0.69
	History	.229**	.176	.342**	0.78
	Education	.285**	.290**	.305**	0.37
	Psychology	.128	.048	.196*	0.71
	Economics	.104	.119	.022	0.32
	Comm Disorders	.183**	.268	.179**	0.20
GRE-V					
	Chemistry	-.110	-.173	-.017	0.57
	Biology	.203*	.253*	.109	0.62
	English	.319**	.294**	.334**	0.30
	History	.226**	.268**	.162	0.48
	Education	.296**	.299**	.299**	0.00
	Psychology	.090	.100	.033	0.31
	Economics	-.032	-.041	-.068	0.08
	Comm Disorders	.158*	.270	.153*	0.26
GRE-Q					
	Chemistry	.117	.160	.037	0.48
	Biology	.245**	.224	.311**	0.40
	English	.146**	.098	.188**	0.58
	History	.165*	.027	.442**	1.95
	Education	.196**	.205**	.226**	0.49
	Psychology	.108	-.030	.294**	1.58
	Economics	.216*	.230*	.168	0.22
	Comm Disorders	.142*	.224	.138*	0.15

* $p < .05$ ** $p < .01$

Table 3

Gender Differences in Prediction of Cumulative GPA From GRE Scores

		Mean Error				
Predictor	Department	Male	Female	*F*	*df*	*p*
GRE-V&Q						
	Chemistry	0.0200	-0.0374	0.62	1, 150	.4329
	Biology	0.0505	-0.0461	1.97	1, 151	.1630
	English	0.0237	-0.0160	1.04	1, 365	.3083
	History	0.0274	-0.0478	2.49	1, 171	.1162
	Education	0.0672	-0.0318	102.54	1, 4583	.0001
	Psychology	0.1370	-0.0758	12.80	1, 217	.0004
	Economics	0.0114	-0.0345	0.17	1, 135	.6810
	Comm Disorders	0.0744	-0.0045	0.62	1, 227	.4337
GRE-V						
	Chemistry	0.0226	-0.0422	0.80	1, 150	.3727
	Biology	0.0419	-0.0383	1.29	1, 151	.2576
	English	0.0130	-0.0088	0.33	1, 365	.5689
	History	0.0266	-0.0465	2.35	1, 171	.1274
	Education	0.0543	-0.0257	66.89	1, 4583	.0001
	Psychology	0.1293	-0.0716	11.23	1, 217	.0010
	Economics	0.0196	-0.0593	0.50	1, 135	.4818
	Comm Disorders	0.0782	-0.0047	0.67	1, 227	.4129
GRE-Q						
	Chemistry	0.0196	-0.0366	0.60	1, 150	.4398
	Biology	0.0681	-0.0621	3.53	1, 151	.0624
	English	0.0260	-0.0176	1.19	1, 365	.2760
	History	0.0281	-0.0490	2.55	1, 171	.1124
	Education	0.0683	-0.0323	101.13	1, 4583	.0001
	Psychology	0.1473	-0.8148	14.85	1, 217	.0002
	Economics	0.0193	-0.0584	0.51	1, 135	.4781
	Comm Disorders	0.0709	-0.0043	0.55	1, 227	.4588